

ATTACHMENT H

TECHNICAL MEMORANDUM,
GEOMORPH™ SAP STATISTICAL CALIBRATION AND
VERIFICATION

TECHNICAL MEMORANDUM

DATE: June 20, 2006

TO: Al Taylor, Michigan Department of Environmental Quality
Greg Rudloff, United States Environmental Protection Agency, Region 5

FROM: Jim Tolbert, Senior Scientist, Earth Tech, Inc.
John Wolfe, Senior Scientist, LimnoTech, Inc.
Peter Simon, Project Manager, Ann Arbor Technical Services, Inc.

SUBJECT: ***GeoMorph*[™] Sampling and Analysis Plan Statistical Calibration and Verification
Upper Tittabawassee River, Midland, Michigan**

PURPOSE

As indicated in the *GeoMorph*[™] Sampling and Analysis Plan for the Upper Tittabawassee River (ATS 2006; SAP), Ann Arbor Technical Services, Inc. (ATS) has developed a statistical protocol for calibration and verification of the *GeoMorph*[™] sampling design proposed for the Upper Tittabawassee River (UTR). This technical memorandum details the statistical procedures and presents the calibration approach that will be used as part of the UTR *GeoMorph*[™] SAP to assess the adequacy of site characterization. The following sections outline the statistical calibration and verification approach that will be used throughout the implementation of the SAP.

BACKGROUND

A principle of the *GeoMorph*[™] sampling design is that there is correlation between the variation in furans and dioxin concentrations in the floodway soils that are associated with distinct fluvial deposition areas. This is based on the relationship between fluvial systems and sediment deposition that are best characterized through geomorphological principles and fluvial processes. The *GeoMorph*[™] sampling design is based on collecting representative soil sampling from distinct fluvial geomorphic features to characterize the furan and dioxin concentrations associated with the soils from these geomorphic features.

The Michigan Department of Environmental Quality (MDEQ) recognizes that biased and random sampling strategies may be utilized as the basis for a sampling design. In some cases random sampling is preferred, while in other cases where extensive knowledge of site conditions is available, biased sampling provides the most accurate characterization of an environmental media.

MDEQ has expressed a desire to statistically calibrate and verify the *GeoMorph*[™] sample design by comparing the efficiencies and results generated by the *GeoMorph*[™] process to that of random sampling designs. In general, standard statistical protocols will be applied to select the most appropriate statistical test(s) based on evaluations of the data set(s). The evaluations will consider a number of variables including; number of data values, satisfaction of normal distribution model assumptions (independence, normal distribution), and potential influence of outliers on data analyses to select the most appropriate parametric and/or nonparametric statistical tests. Overall, four separate approaches will be used to support the *GeoMorph*[™] sampling design.

- The first procedure evaluates the assumption that part of the variation in the concentrations of furans and dioxins in floodway soils can be explained by accounting for fluvial deposition areas in the floodway.
- The second procedure will be utilized to evaluate outliers in furan and dioxin concentrations within geomorphic features during the field activities so that these outliers can be addressed in a consistent and timely manner.
- The third and fourth procedures compare the biased *GeoMorph*[™] sampling design to random sampling designs.

STATISTICAL PROCEDURES

1. EVALUATE *GeoMorph*[™] STRATIFICATION

Through the course of the proposed work, soil borings will be set at hundreds of sample locations and thousands of soil samples will be analyzed for furans and dioxins. Each of these sample locations will be specifically tied to a geomorphic feature as well as other aspects that describe the sample location such as a distance from the river bank, a distance from Midland, and an elevation. Initially, an analysis of variance (ANOVA) will be performed to determine whether geomorphic features explain a significant amount of the variation in the furan and dioxin concentrations. The S3TM (MDEQ, 2002; S3TM p. 4.25) supports the use of ANOVA to evaluate if “*analytical results differ significantly among strata*”, and whether use of the strata are “*necessary and statistically valid*.”

ANOVA will be conducted on a subset of furan and dioxin congener concentration data of greatest quantitative significance for characterizing extent and distribution of impacts and for risk assessment. The distribution of concentrations within each *GeoMorph*[™] stratum will be evaluated for consistency with normal or lognormal distributions, to determine whether it is more statistically appropriate to perform ANOVA on concentrations or the logarithm of concentrations. The prevalence of non-detects (censored data) and outliers will also be addressed, as described below in this section. If the data set(s) do not support the assumption of a normal or lognormal distribution, then a nonparametric statistical comparison may be utilized to evaluate the variation in the furan and dioxin concentrations.

1.1 DATA SET

The data set will consist of all of the soil sample locations where soil samples are analyzed for furans and dioxins. The data set will include categorical variables that are composed of a limited set of categories (e.g., mapped geomorphic feature type and reach of the river) and numeric variables that are assumed to be from a continuous numerical domain (e.g., concentrations, distances from river, and depth of sample). At most sample locations soil samples will be analyzed at multiple depths within the core and these analyses will extend down into un-impacted materials below the sediments that were deposited before the period of interest. It is anticipated that analytical data from the depth containing the maximum detected concentration will be used to represent the concentration at a given sample location. This method will focus the ANOVA on explaining the variation in the most impacted soil horizon within a sample location

1.2 CENSORED DATA

In some samples, it is anticipated that the concentration will be less than the detection limit for the analysis. When this occurs, the data point will still be utilized in this comparison by substituting a value of one half of the detection limit. This procedure reduces the variation in concentration. For this reason, if the proportion of censored data is greater than 50 percent, other modifications may be utilized (e.g. the

data set may also be evaluated with only the samples that contained detectable furans and dioxins or the maximum concentration over all depths at a sample location may be used). It is anticipated that more than 50 percent of the sample locations will contain detectable concentrations of furans and dioxins. If greater than 50 percent of the data does not contain detectable furans and dioxins, then it will be assumed that a significant portion of the sample locations have not been impacted by the site, and the data set will be censored down to a subset that contains a greater ratio of detectable concentrations of furans and dioxins to evaluate the variance within samples that may be effected by the deposition from the Tittabawassee River. Any censoring (which is not anticipated) will be reviewed with the MDEQ.

1.3 METHOD

ANOVA analyzes the variation in a variable of interest (in this case furan and dioxin concentrations) and attempts to assign portions of the variation to each member of a set of explanatory variables (in this case geomorphic location indicators), leaving a remainder that is associated with a random or unexplained component. With ANOVA, it is possible to test the null hypothesis of “no effect” of the explanatory variables, where the alternative hypothesis is that there is a relationship between concentration and at least one of the explanatory variables. The hypothesis is tested by constructing an F statistic from the explained and unexplained portions of variation in concentration. If the data set does not support the assumption of a normal or lognormal distribution, then a nonparametric statistical comparison may be utilized to evaluate the variation in the furan and dioxin concentrations. Commercially available software will be used to perform the statistical tests (e.g., STATISTICA, StatPac, statistiXL (an add-in for Excel), or SPSS).

If all explanatory variables are categorical, and the hypothesis of no effect is rejected, then the stratified subsamples can be treated as samples from separate populations, computing separate summary statistics including 95 percent upper confidence limits (UCLs) for each by standard methods. Explanatory variables may also include a combination of categorical variables (e.g., location of sample in a geomorphic feature or vertical stratum) and continuous variables (e.g., distance from river along flow paths or depth below grade). To quantify the effects of each of these explanatory variables, the ANOVA can be structured as a multiple linear regression of concentration on the categorical indicators (i.e., taking on values of 0 or 1) and continuous explanatory variables suspected to have an effect on concentration. Estimated coefficients of the categorical variables then provide estimates of the differences in concentration between *GeoMorph*[™] features and coefficients of continuous variables likewise estimate the impacts of those continuous variables. A continuous variable of particular interest is distance from river along a streamline. The null hypothesis of “no effect” for an individual variable can be tested by constructing a t-statistic, from the estimated coefficient and its standard error. If the hypothesis of “no effect” is rejected, the explanatory variable is retained.

If a mix of categorical and continuous variables is retained, a mean predicted concentration and its distribution can be obtained from the regression results for any given set of explanatory variables (e.g., specifying a particular *GeoMorph*[™] feature and distance from river along a flow path). The method for estimating confidence limits for the mean, including the 95 percent UCL, given specified values of the explanatory variables, is discussed by Neter, Wasserman, and Kutner (p 246, 1990) and other textbooks on linear regression models. If the UCLs are compared to identified criteria then an evaluation of the sample size (n) will be performed. The sample size evaluation will follow the procedures documented in Section 3.2 of the S3TM (MDEQ, 2002) or other available references for statistical procedures for evaluation of sample sizes.

2. SCREENING FOR OUTLIERS IN GEOMORPHIC FEATURES

Abnormally high and low concentrations of furans and dioxins in a geomorphic feature will be identified and addressed during the field activities. Identifying and addressing these outliers early and during the active field period will improve the quality and coverage of the final data set generated on the UTR project.

Initially, the furan and dioxin concentrations from each geomorphic surface within a given reach will be screened for outliers. A graphical evaluation of each population (i.e., each geomorphic feature and similarly mapped geomorphic features within a given reach) will be performed to look for potential outliers and to evaluate the assumptions about the distribution of the data required for ANOVA and required for the outlier test described below. The data and log transformed data will be separately represented in box plots and probability plots. The distribution of data in each population will be evaluated from these plots. Each data set will be evaluated for both normal and lognormal distributions. The results of these evaluations will be compared to determine which distribution provides a better fit to the data.

After initial screening, a formal test will be performed if the presence of one or more outliers is suspected following the procedures in the S3TM (MDEQ, 2002; Section 2.1.4). If only one outlier is suspected based upon the graphical representation, then Grubbs' test will be utilized. If multiple outliers are possible based upon the graphical representations, then Dixon's Test will be used if the sample size is equal to or less than 25, and Rosner's test will be utilized for sample sizes larger than 25.

These tests will be utilized when there are at least four detectable concentrations in the population. If there are less than four detectable concentrations in the population, only graphical methods will be used to evaluate possible outliers. Censored data (not detectable concentrations) will be addressed using the procedures presented in the section on ANOVA. Prior to performing Dixon's Test, the data distribution will be evaluated graphically. If the data are assumed to be normally distributed, then the un-transformed data will be utilized. If the data are assumed to be log normally distributed, then the data will be log transformed prior to use in tests.

All data that are identified as outliers will be evaluated. S3TM (MDEQ, 2002; S3TM p 7.42) provides the following partial list of possible causes for outliers:

- errors in sampling, laboratory analysis, data entry, or transcription;
- an accurate result sampled from a different population;
- an accurate but extreme value from the original population;
- an accurate value that appears to be extreme because of failure to obtain a representative sample, due to insufficient number of samples or biased sampling.

In consideration of the need to evaluate each outlier or set of outliers identified, one of the following actions will be taken.

1. Laboratory records will be reviewed for possible errors in analysis, data entry, or transcription.
2. The geomorphic feature will be evaluated to determine if field data support the conclusion that there may be more than one feature present. If there is supporting evidence to re-classify part of the geomorphic feature, then the geomorphic feature may be re-defined or split into multiple features to address the data generated from the study.
3. The outlier may appear to accurately represent the variation within a geomorphic feature, and the outlier may be accepted as part of the geomorphic feature.

4. Additional soil samples may be collected and analyzed to further evaluate and delineate the outlier within a mapped geomorphic feature.

Performing the outlier evaluation at a 95 percent level of confidence (i.e., $\alpha = 0.05$), it is expected that one out of every 20 comparisons will identify an outlier even if there are no outliers present in any data set. Furthermore, when testing for multiple outliers (potentially three outliers evaluated in every data set), the potential frequency of falsely identifying outliers increases. There are 15 reaches being evaluated, and there are typically more than six features in each reach. Therefore, it is expected that over 100 comparisons will be made. With this number of comparisons, it is expected that multiple tests will falsely identify an outlier when one does not exist. In addition, the outlier test assumes normality. With the small data sets that will be tested, we will not be able to provide a statistically robust test of this assumption. Some populations will likely not be normal or lognormal. The expectation that distributional assumptions will be incorrect in some cases increases the probability of falsely identifying an outlier when no outlier actually is present. Therefore, not all outliers will be eliminated from the data set, but each outlier will be reviewed with the MDEQ, and addressed in the final report. The outlier evaluation is included to screen the rapid turnaround analytical data and to focus additional sampling and field mapping activities within the 2006 *GeoMorph*[™] sampling program.

3. COMPARISON TO RANDOM-ON-GRID SAMPLE DESIGN METHOD

3.1 OVERVIEW

The objective of this evaluation is to compare the data generated from a *GeoMorph*[™] biased transect sample design, to the data generated from a random-on-grid sample design, an alternative sampling strategy discussed in S3TM (MDEQ, 2002; S3TM p 4.24). For this purpose, an existing random-on-grid dataset will be employed. It is proposed that data from “Scoping Study” Area 1 (CH2M Hill, 2005) be used for this comparison. The comparison will examine differences in uncertainty bounds of mean predicted concentrations using the two strategies. This will include a comparison of 95 percent UCLs of means, as well as lower confidence limits of mean predictions, and will be supported by a comparison of other summary statistics generated by the alternative strategies.

3.2 DATA SETS

Data from the “Scoping Study” Area 1 (CH2M Hill, 2005) will be used for this comparison. Surface soil concentrations from both datasets will be used for the comparison, to ensure direct comparability. The same congeners analyzed in the Evaluation of *GeoMorph*[™] Stratification will be used in this comparison, if available in the random-on-grid dataset.

3.2.1 *GeoMorph* Biased Transect Sample Data Set

A subset of the samples generated through the implementation of the SAP will form the data set for the *GeoMorph*[™] Transect Sampling. The subset will consist of the samples from Reaches N and O only; this subset of samples will represent the soil conditions at the “Scoping Study” Area 1 and will be directly comparable to the random on grid sample data set. When a boring is collected, the surface soil samples will be used to allow direct comparison to the surface soil samples collected during the random-on-grid sampling. If there is more than one sample interval collected across the top 6 inches of soil, the maximum concentration observed in any sample covering the top 6 inches of soil will be used to represent the sample location.

3.2.2 *Random-on-Grid Sample Set*

The “Scoping Study” Area 1 data set will be utilized as a single data set from the random on grid sampling approach. The sample size will exceed 50 samples. The sample locations from Area 1 are shown on Figure 1.

3.2.3 *Censored Data and Screening for Outliers*

Treatment of censored data and evaluation of outliers will follow the procedures described in Sections 1 above for the Evaluation of *GeoMorph*[™] Stratification.

3.3 SPECIFIC OBJECTIVES – RANDOM-ON-GRID SAMPLE DESIGN COMPARISON

The objective of this procedure is to compare the confidence intervals for mean predicted concentration using the random-on-grid sample design with confidence intervals for mean predicted concentrations using *GeoMorph*[™] surfaces, based on the *GeoMorph*[™] biased transect sampling design. Compare standard errors of estimates of means, and 95 percent UCLs and lower confidence limits (LCLs) of means, with random-on-grid sample design and stratified design. Fixed interval transect data set(s) and the existing data set(s) from “Scoping Study” Area 1 will be evaluated to determine if an additional mapped geomorphic feature needs to be sampled with a random-on-grid sample design to evaluate adequacy geomorphic feature characterization. If required, the need and location of additional random-on-grid sampling will be reviewed with MDEQ as part of the iterative process.

3.4 METHOD – COMPARATIVE STATISTICAL COMPARISONS

The two pooled data sets and the stratified data sets from the *GeoMorph*[™] features will first be compared using standard statistical methods that include creating graphical comparative statistical outputs to compare the populations:

- Box plots of the data and of log transformed data;
- Probability plots of the data and of log transformed data;

Comparative statistics on the two populations will be evaluated side by side:

- Sample size (n)
- Number and percent of censored data (concentrations reported as not detectable)
- Mean (x)
- Median
- Every 10th percentile to assess the shape of the distribution
- Standard deviation (s)
- Skewness

Determine whether data distributions are approximately normal, lognormal, or neither, based on visual inspection of probability plots taking into account outliers, and hypothesis tests in S3TM. To evaluate the gain in sampling efficiency from *GeoMorph*[™] stratification, compute sample means, standard errors of means, and 95 percent upper and lower confidence limits of means for

- pooled data from random-on-grid sample;
- the unstratified *GeoMorph*[™] sample; and
- individual *GeoMorph*[™] polygons.

3.5 ABILITY TO MEET PROJECT OBJECTIVES

The objective of the *GeoMorph*[™] investigation is to adequately characterize the distribution of furans and dioxins in the Upper Tittabawassee River and associated floodplain, maximizing information obtained for a given sampling effort. The comparison of uncertainty bounds for different sampling approaches will be utilized to review this overall project objective.

4. COMPARISON TO FIXED INTERVAL TRANSECT SAMPLE DESIGN METHOD

4.1 OVERVIEW

The objective of this evaluation is to compare the data generated from a *GeoMorph*[™] biased transect sample design, to the data generated from a fixed interval transect sample design (i.e., MDEQ Validation Transect Design). This evaluation will compare the data generated on selected validation transects, where a tighter boring interval is maintained, to the data generated with the *GeoMorph*[™] biased transect sampling design through the same reach. The data will be used to evaluate the measured concentrations of furans and dioxins within the floodway soils using the fixed interval transect approach versus a *GeoMorph*[™] biased transect sampling approach.

4.2 DATA SET

4.2.1 GeoMorph[™] Biased Transect Sample Data Set

The data set for the *GeoMorph*[™] Transect Data will consist of a subset of the samples generated through the implementation of the SAP will form the data set. The subset of samples will consist of the surface samples from the sample locations contained within the same reach as the validation transect. This subset of samples will represent the variability of the data documented by the *GeoMorph*[™] sample design within the same area as the validation transect described below.

4.2.2 Validation Transect Data Set

Initially, samples will be collected from one validation transect on a fixed interval with distance along the transect and also with distance away from the transect along perpendicular extensions from the transect. The transect has been pre-selected and is shown in Figure 2. The transect will consist of sample locations every 100 feet moving away from the Tittabawassee River in one direction and will include three extensions from the main transect that extend away at a right angle from the transect. Each perpendicular extension will be sampled at 100 foot intervals extending away from the main transect in each direction (10 total sample locations on the transect). The perpendicular extensions will be placed each 500 feet along the main transect (at the first sample location, the sixth sample location, and the eleventh sample location moving away from the river). A surface soil sample will be collected from each of these sample locations that represents the upper horizon of the soil. A second fixed interval validation transect may be implemented as part of the iterative sampling process. The need and location of the second transect will be reviewed with MDEQ.

The data from the validation transect will form a data set. To create the most cost-effective sampling program, some of the data points on the validation transect will be part of both the *GeoMorph*[™]-based transect sample data set and also the MDEQ fixed interval validation transect data set. Therefore, these two populations will not be entirely independent but the two populations are expected to provide a basis for comparison.

4.2.3 Censored Data and Screening for Outliers

Treatment of censored data and evaluation of outliers will follow the procedures described in Section 1 for the Evaluation of *GeoMorph*[™] Stratification.

4.3 SPECIFIC OBJECTIVE - FIXED INTERVAL TRANSECT SAMPLE DESIGN COMPARISON

The objective of this procedure is to evaluate the concentrations of furans and dioxins within the floodway soils using the fixed interval transect approach versus a *GeoMorph*[™] biased transect sampling approach and to test the hypothesis that geomorphological stratification improves sampling efficiency.

4.4 METHOD – COMPARATIVE STATISTICAL COMPARISONS

The validation dataset will be evaluated as a whole and as component transect datasets. Similarly, the *GeoMorph*[™] dataset will be evaluated as a whole and as component datasets for individual *GeoMorph*[™] surfaces. The pooled and component data sets will first be compared using standard statistical methods that will include creating graphical comparative statistical outputs to compare the populations:

- Box plots of the data and of log transformed data;
- Probability plots of the data and of log transformed data.

Comparative statistics on the two populations will be evaluated side by side:

- Sample size (n)
- Number and percent of censored data (concentrations reported as not detectable)
- Mean (x)
- Median
- Every 10th percentile to assess the shape of the distribution
- Standard deviation (s)
- Skewness

Determine whether data distributions are approximately normal, lognormal, or neither, based on visual inspection of probability plots taking into account outliers, and hypothesis tests in S3TM. To evaluate the gain in sampling efficiency from *GeoMorph*[™] stratification, compute sample means, standard errors of means, and 95 percent upper and lower confidence limits of means for:

- pooled data from validation transects;
- the individual component cross-sections aligned from river to upland;
- the pooled (unstratified) *GeoMorph*[™] sample; and
- individual *GeoMorph*[™] polygons.

4.5 ABILITY TO MEET PROJECT OBJECTIVES

The objective of the *GeoMorph*[™] investigation is to adequately characterize the distribution of furans and dioxins in the Upper Tittabawassee River and associated floodplain, maximizing information obtained for a given sampling effort. The comparison of uncertainty bounds for different sampling approaches will be utilized to review this overall project objective.

REFERENCES

Ann Arbor Technical Services, Inc. (ATS). 2006. *GeoMorph*[™] *Sampling and Analysis Plan, Upper Tittabawassee River*. June.

CH2M Hill. 2005. *Tittabawassee River Floodplain Scoping Study Work Plan-Revised*. July.

Michigan Department of Environmental Quality, Remediation and Redevelopment Division. (MDEQ). 2002. *Sampling Strategies and Statistics Training Manual*.

Neter, John, William Wasserman, and Michael H. Kutner. 1990. *Applied Linear Statistical Models*. Irwin. Burr Ridge, Illinois.



